

Application No. 10/626,875
Amd. dated October 23, 2006
Reply to July 2006 Communication

Docket No. D/A3053-311291

Listing of Claims:

1. (Currently Amended) A computer-implemented method of determining predictive models for a linked event detection system comprising the steps of:
 - determining source-identified training stories;
 - determining inter-story similarity vectors in a memory for at least one story-pair of the source-identified training stories;
 - determining link label information for the at least one story-pair;
 - determining and storing at least one predictive model in the memory based on the inter-story similarity vectors and the link label information; ~~and~~
 - ~~indicating a link between other story-pairs based on the predictive model and the inter-story similarity vector.~~
2. (Original) The method of claim 1, wherein the step of determining inter-story similarity vectors comprises the steps of:
 - determining at least one inter-story similarity metric for the story-pairs; and
 - determining at least one source-pair statistics for the at least one story-pair.
3. (Original) The method of claim 2, wherein determining inter-story similarity vectors further comprise the step of normalizing the inter-story similarity metric based on the source-pair statistics.
4. (Original) The method of claim 2, wherein determining inter-story similarity vectors further comprise the step of incrementally normalizing the inter-story similarity metric based on the source-pair statistics.

Application No. 10/626,875
Amd. dated October 23, 2006
Reply to July 2006 Communication

Docket No. D/A3053-311291

5. (Original) The method of claim 2, wherein the inter-story similarity metric is normalized based on at least one of subtraction and division.
6. (Original) The method of claim 2, wherein the inter-story similarity metric is at least one of a probability based similarity metric and a Euclidean based similarity metric.
7. (Original) The method of claim 6, wherein the probability based inter-story similarity metric is at least one of a Hellinger, a Tanimoto and a clarity distance based metric.
8. (Original) The method of claim 6, wherein the Euclidean based inter-story similarity metric is a cosine-distance based metric.
9. (Original) The method of claim 1, further comprising the step of transforming the
source-identified training stories.
10. (Original) The method of claim 9, wherein transforming the source-identified training stories is at least one of translating, transcribing and linguistically transforming.
11. (Previously Amended) The method of claim 2, wherein the inter-story similarity metrics are based on terms in at least one source-identified term frequency-inverse story frequency models.
12. (Original) The method of claim 11, wherein the terms in source-identified term frequency-inverse story frequency models are based on language.
13. (Original) The method of claim 11, wherein determining terms comprises the steps:
determining a reference language; and
determining reference language and non-reference language terms.

Application No. 10/626,875
Amd. dated October 23, 2006
Reply to July 2006 Communication

Docket No. D/A3053-311291

14. (Original) The method of claim 2, wherein the at least one inter-story similarity metric is normalized based on at least one of a source-pair identified similarity statistic.
15. (Original) The method of claim 1, wherein the at least one predictive model is at least one of: a classifier, a support vector machine, a decision tree and a Naive-Bayes classifier.
16. (Original) The method of claim 2, wherein at least one of the source-pair similarity statistics are determined based on a source hierarchy.
17. (Original) The method of claim 16 wherein the source hierarchy is determined based on at least one source characteristic.
18. (Original) The method of claim 16 wherein the source characteristic is at least one of a language characteristic, an input mode characteristic, a genre characteristic, a source name characteristic and a transformation characteristic.
19. (Original) The method of claim 16 wherein the source-pair similarity statistic for a new source is determined based on at least one source characteristic of the new source.
20. (Currently Amended) A linked event detection training system comprising:
 - an input/output circuit;
 - a memory;
 - a processor that receives source-identified training stories and associated link label information for at least one story-pair via the input/output circuit;
 - an inter-story similarity vector determining circuit that determines an inter-story similarity vectors in the memory for at least one story-pair of the source-identified training stories; and

Application No. 10/626,875
Amd. dated October 23, 2006
Reply to July 2006 Communication

Docket No. D/A3053-311291

a predictive model determining circuit that determines and stores at least one predictive model in the memory based on the inter-story similarity vectors and the link label ~~information and which indicates a link between other story-pairs based on the predictive model and the inter-story similarity vector.~~

21. (Original) The system of claim 20, wherein the inter-story similarity vector determining circuit is comprised of:

a similarity metric determining circuit that determines at least one inter-story similarity metric for the at least one story-pair; and

a similarity statistics determining circuit that determines at least one source-pair statistic for the at least one story-pair.

22. (Original) The system of claim 21, wherein the inter-story similarity vector determining circuit normalizes the inter-story similarity metric based on the source-pair statistics.

23. (Original) The system of claim 21, wherein the inter-story similarity vector determining circuit incrementally normalizes the inter-story similarity metric based on the source-pair statistics.

24. (Original) The system of claim 21, wherein at least one of the inter-story similarity metrics is normalized based on at least one of a subtraction and a division operation.

25. (Original) The system of claim 21, wherein at least one of the inter-story similarity metrics is at least one of a probability based similarity metric and a Euclidean based similarity metric.

26. (Original) The system of claim 25, wherein the probability based inter-story similarity metric is at least one of a Hellinger, a Tanimoto and a clarity distance based metric.

Application No. 10/626,875
Amd. dated October 23, 2006
Reply to July 2006 Communication

Docket No. D/A3053-311291

27. (Original) The system of claim 25, wherein the Euclidean based inter-story similarity metric is a cosine-distance based metric.
28. (Original) The system of claim 20, wherein the source-identified training stories are transformed.
29. (Original) The system of claim 28, wherein transforming the source-identified training stories is at least one of translating, transcribing and linguistically transforming.
30. (Currently Amended) The system of ~~claim 21~~ claim 21, wherein the inter-story similarity metrics are based on terms in at least one source-identified term frequency-inverse story frequency model.
31. (Original) The system of claim 30, wherein the terms in the source-identified term frequency-inverse story frequency models are based on language.
32. (Original) The system of claim 30, wherein the processor determines terms based on a reference language; and determining reference language and non-reference language terms.
33. (Original) The system of claim 21 wherein the at least one inter-story similarity metric is normalized based on at least one of a source-pair identified similarity statistic.
34. (Original) The system of claim 20, wherein the at least one predictive model is at least one of: a classifier, a support vector machine, a decision tree and a Naive-Bayes classifier.
35. (Original) The system of claim 21, wherein the source-pair identified similarity statistic is determined based on a source hierarchy.

Application No. 10/626,875
Amd. dated October 23, 2006
Reply to July 2006 Communication

Docket No. D/A3053-311291

36. (Original) The system of claim 35, wherein the source hierarchy is determined based on at least one of a source characteristic.
37. (Original) The system of claim 35, wherein the source characteristic is at least one of a language characteristic, an input mode characteristic, a genre characteristic, a source name characteristic and a transformation characteristic.
38. (Original) The system of claim 35, wherein the source-pair similarity statistic for a new source is determined based on at least one source characteristics of the new source.
39. (Currently Amended) A computer-implemented method of linked event detection comprising the steps of:
- determining source-identified ~~training~~-stories;
 - determining inter-story similarity vectors in a memory for the story-pairs of the source-identified stories;
 - determining at least one predictive model in a memory for link detection; and
 - determining a link between the story-pairs based on the predictive model and the inter-story similarity vectors; and
 - indicating the link.
40. (Previously Amended) The method of claim 39, wherein the step of determining inter-story similarity vectors comprises the steps of:
- determining at least one inter-story similarity metric for each story-pair; and
 - determining source-pair statistics for the story-pairs.

Application No. 10/626,875
Amd. dated October 23, 2006
Reply to July 2006 Communication

Docket No. D/A3053-311291

41. (Original) The method of claim 40, wherein determining inter-story similarity vectors further comprise the step of normalizing the inter-story similarity metric based on the source-pair statistics.
42. (Original) The method of claim 40, wherein determining inter-story similarity vectors further comprise the step of incrementally normalizing the inter-story similarity metric based on the source-pair statistics.
43. (Original) The method of claim 40, wherein the inter-story similarity metric is normalized based on at least one of subtraction and division.
44. (Original) The method of claim 40, wherein the inter-story similarity metric is at least one of a probability based similarity metric and a Euclidean based similarity metric.
45. (Original) The method of claim 44, wherein the probability based inter-story similarity metric is at least one of a Hellinger, a Tanimoto and a clarity distance based metric.
46. (Original) The method of claim 44, wherein the Euclidean based similarity metric is a cosine-distance based metric.
47. (Original) The method of claim 39, further comprising the step of transforming the source-identified training stories.
48. (Original) The method of claim 47, wherein transforming the source-identified training stories is at least one of translating, transcribing and linguistically transforming.
49. (Currently Amended) The method of ~~claim 40~~claim 40, wherein the inter-story similarity metrics are based on terms in at least one source-identified term frequency-inverse story frequency models.

Application No. 10/626,875
Amd. dated October 23, 2006
Reply to July 2006 Communication

Docket No. D/A3053-311291

50. (Original) The method of claim 49, wherein the terms in source-identified term frequency-inverse story frequency models are based on language.

51. (Original) The method of claim 49, wherein determining terms comprises the steps:

determining a reference language; and

determining reference language and non-reference language terms.

52. (Original) The method of claim 40, wherein the at least one inter-story similarity metric is normalized based on at least one of a source-pair identified similarity statistic.

53. (Original) The method of claim 39, wherein the at least one predictive model is at least one of: a classifier, a support vector machine and a decision tree, a Naive-Bayes-classifier.

54. (Original) The method of claim 40, wherein the source-pair identified similarity statistic is determined based on a source hierarchy.

55. (Original) The method of claim 54, wherein the source hierarchy is determined based on at least one of a source characteristic.

56. (Original) The method of claim 54, wherein the source characteristic is at least one of a language characteristic, an input mode characteristic, a genre characteristic, a source name characteristic and a transformation characteristic.

57. (Original) The method of claim 54, wherein the source-pair similarity statistic for a new source is determined based on at least one source characteristics of the new source.

Application No. 10/626,875
Amd. dated October 23, 2006
Reply to July 2006 Communication

Docket No. D/A3053-311291

58. (Currently Amended) A linked event detection system comprising:

- an input/output circuit;
- a memory;
- a processor that receives source-identified ~~training~~-stories via the input/output circuit;
- an inter-story similarity vector determining circuit that determines inter-story similarity vectors in the memory for the story-pairs of the source-identified stories; and
- a link determining circuit that determines and indicates links between story-pairs based on a predictive model in the memory and the inter-story similarity vectors.

59. (Original) The method of claim 58, wherein the inter-story similarity vector determining circuit is comprised of:

- a similarity metric determining circuit that determines at least one inter-story similarity metric for the story-pairs; and
- a similarity statistics determining circuit that determines source-pair statistics for the story-pairs.

60. (Original) The system of claim 59, wherein the inter-story similarity vector determining circuit normalizes the inter-story similarity metric based on the source-pair statistics.

61. (Original) The system of claim 59, wherein the inter-story similarity vector determining circuit incrementally normalizes the inter-story similarity metric based on the source-pair statistics.

62. (Original) The system of claim 59, wherein at least one of the inter-story similarity metrics is normalized based on at least one of a subtraction and a division operation.

Application No. 10/626,875
Amd. dated October 23, 2006
Reply to July 2006 Communication

Docket No. D/A3053-311291

63. (Original) The system of claim 59, wherein at least one of the inter-story similarity metrics is at least one of a probability based similarity metric and a Euclidean based similarity metric.
64. (Original) The system of claim 63, wherein the probability based inter-story similarity metric is at least one of a Hellinger, a Tanimoto and a clarity distance based metric.
65. (Original) The system of claim 63, wherein the Euclidean based inter-story similarity metric is a cosine-distance based metric.
66. (Original) The system of claim 58, wherein the source-identified training stories are transformed.
67. (Original) The system of claim 66, wherein transforming the source-identified training stories is at least one of translating, transcribing and linguistically transforming.
68. (Previously Amended) The system of claim 59, wherein the inter-story similarity metrics are based on terms in at least one source-identified term frequency-inverse story frequency model.
69. (Original) The system of claim 68, wherein the terms in the source-identified term frequency-inverse story frequency models are based on language.
70. (Original) The system of claim 68, wherein the processor determines terms based on a reference language; and non-reference language terms.
71. (Original) The system of claim 59, wherein the at least one inter-story similarity metric is normalized based on at least one of a source-pair identified similarity statistic.

Application No. 10/626,875
Amd. dated October 23, 2006
Reply to July 2006 Communication

Docket No. D/A3053-311291

72. (Original) The system of claim 58, wherein the predictive model is at least one of: a classifier, a support vector machine and a decision tree, a Naive-Bayes classifier.
73. (Original) The system of claim 59, wherein the source-pair identified similarity statistic is determined based on a source hierarchy.
74. (Original) The system of claim 73, wherein the source hierarchy is determined based on at least one of a source characteristic.
75. (Original) The system of claim 73, wherein the source characteristic is at least one of a language characteristic, an input mode characteristic, a genre characteristic, a source name characteristic and a transformation characteristic.
76. (Original) The system of claim 73, wherein the source-pair similarity statistic for a new source is determined based on at least one source characteristics of the new source.
77. (Currently Amended) A method of determining a stopword list comprising the steps of:
- determining a source-identified training corpus of text information;
 - determining a verified first source-mode transformation of the source-identified training corpus text from a first source mode to a second source mode;
 - determining an un-verified second source-mode transformation of the source-identified training corpus text from a first source mode to a second source mode;

Application No. 10/626,875
Amd. dated October 23, 2006
Reply to July 2006 Communication

Docket No. D/A3053-311291

determining at least one transformation errors associated with distribution differences between the first and second transformations and identified sources;

determining and storing at least one source-specific transformation actions for the determined transformation errors in a memory; and

identifying and transforming transformation errors in other transformed source-identified texts based on the source-specific transformation actions in the memory.

78. (Original) The method of claim 77, wherein the first source mode is at least one of a text source, an optical character recognition source and an automatic speech recognition source.

79. (Original) The method of claim 77, wherein the second source mode is at least one of a text source, an optical character recognition source and an automatic speech recognition source.

80. (Original) The method of claim 77, wherein the source-specific transformation is at least one of a removal, a repair and a normalization transformation.

81. (Currently Amended) Computer readable storage medium comprising: computer readable program code embodied on the computer readable storage medium, the computer readable program code usable-executable to program a computer to determine at least one predictive model for a linked event detection system comprising the steps of:

determining source-identified training stories;

determining inter-story similarity vectors in a memory for at least one story-pair of the source-identified training stories;

Application No. 10/626,875
Amd. dated October 23, 2006
Reply to July 2006 Communication

Docket No. D/A3053-311291

determining link label information for the at least one story-pair; and

determining at least one predictive model in the memory based on the inter-story similarity vectors and the link label information; ~~and indicating a link between other story pairs based on the predictive model and the inter-story similarity vector.~~

82. (Currently Amended) Computer readable storage medium comprising: computer readable program code embodied on the computer readable storage medium, the computer readable program code usable to program a computer to determine at least one predictive model for a linked event detection system comprising:

~~instructions for determining to determine~~ source-identified training stories;

~~instructions to determine for determining~~ inter-story similarity vectors in a memory for at least one story-pair of the source-identified training stories;

~~instructions to determine for determining~~ link label information for the at least one story-pair; and

~~instructions to determine for determining~~ at least one predictive model in the memory based on the inter-story similarity vectors and the link label information; ~~and~~

~~instructions for indicating a link between other story pairs based on the predictive model and the inter-story similarity vector.~~

83. (Currently Amended) Computer readable storage medium comprising: computer readable program code embodied on the

Application No. 10/626,875
Amd. dated October 23, 2006
Reply to July 2006 Communication

Docket No. D/A3053-311291

computer readable storage medium, the computer readable program code ~~usable-executable~~ to program a computer to detect linked events comprising the steps of:

- determining source-identified ~~training~~-stories;
- determining inter-story similarity vectors in a memory for the at least one story-pair of the source-identified stories;
- determining at least one predictive model in the memory for link detection;
- determining a link between story-pairs based on the at least one predictive model and the inter-story similarity vectors; and
- indicating the link.

84. (Currently Amended) Computer readable storage medium comprising: computer readable program code embodied on the computer readable storage medium, the computer readable program code ~~usable-executable~~ to program a computer to detect linked events comprising the steps of:

- ~~instructions to determine~~for determining source-identified ~~training~~-stories;
- ~~instructions to determine~~for determining inter-story similarity vectors in a memory for the at least one story-pair of the source-identified stories;
- ~~instructions to determine~~for determining at least one predictive model in a memory for link detection;
- ~~instructions to determine~~for determining a link between story-pairs based on the predictive model and the inter-story similarity vectors; and
- instructions to indicate ~~indicating~~ the link.

Application No. 10/626,875
Amd. dated October 23, 2006
Reply to July 2006 Communication

Docket No. D/A3053-311291

85. (Original) The method of claim 2, wherein determining at least one source-pair statistic for the at least one story-pair is based on at least one of a similarity metric and a statistic associated with the metric.
86. (Original) The system of claim 21, wherein determining at least one source-pair statistic for the at least one story-pair is based on at least one of a similarity metric and a statistic associated with the metric.
87. (Original) The method of claim 39, wherein at least one of the predictive models is a trained predictive model.
88. (Original) The system of claim 58, wherein at least one of the predictive models is a trained predictive model.